DOCUMENT RESUME

ED 185 586 CS 205 505

AUTHOR Witte, Stephen P.
TITLE Text Cohesion and the Study of Writter Discourse.
PUB DATE Mar 80
NOTE 17p.: Paper presented at the Annual Meeting of the Conference on College Composition and Communication (31st, Washington, DC, March 13-15, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College English: College Freshmen: Comparative Analysis: *Discourse Analysis: Expository Writing: Higher Education: *Literary Devices: *Writing (Composition): *Writing Skills
IDENTIFIERS *Cohesion (Written Composition): *Writing Research

ABSTRACT

M. A. K. Halliday and R. Hasan have provided a system for analyzing the cohesive relations that enable a sequence of T-units to be considered a complete text. (A T-unit consists of a main clause and all its dependent clauses.) These concepts of cohesion proved effective in analyzing the differences between five "high" and five "low" quality essays written by college freshmen. At the most general level of analysis, the two essay sets differed according to the mean number of words separating cohesive ties, with more cohesive ties appearing in T-units from the good essays. The Halliday-Hasan classifications for span of text also helped distinguish between good and poor writers, with the good writers using substantially more immediate ties and fewer mediated-remote ties than the poor writers. Mediated ties appeared about eight times more frequently in the better essays, a finding that further supported the notion of extending the lexical-semantic domain of a given T-unit beyond its normal bounds. It was proposed that poor writers' dependence on mediated-remote ties partially explained their frequent failure to elaborate and extend lexical-semantic domains, a shortcoming that prevented them from providing as much really new information as the good writers did. (Examples of each of the types of cohesive ties in the Halliday-Hasan system are attached.) (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185586

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TEXT COHESION AND THE STUDY OF WRITTEN DISCOURSE

by

Stephen P. Witte

University of Texas at Austin

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephen P. Witte

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

March, 1980

Paper read during the annual
Conference on College Composition and Communication
March 13-15, 1980
Washington, D.C.

CS205505

Text Cohesion and the Study of Written Discourse

There is no doubt that much of our teaching of writing is concerned with the nature and production of whole discourses, discourses that extend well beyond individual sentences to units consisting of several paragraphs, each linked to the preceding one and to the following one. Nevertheless, those familiar with the corpus of empirical research in written composition produced over the past decade and a half will probably agree that sentence-level syntax has been a dominant, if not the predominant, concern. This concern is not surprising: a great deal of our effort as teachers of writing is focused on the sentence; and many of us who would like to think that we apprehend something of "the Pythagorean power by which number holds sway above the flux" of the universe of discourse and discoursing have been more than willing to steal from Chomsky's (1957) comet fire for our small tapers and to follow the leads of Hunt (1965) and Mellon (1969) in bringing light to our otherwise dark worlds. Now, however, the light of syntactic analyses seems less illuminating, principally because such studies have failed to reveal anything about the text relationships that obtain across sentence boundaries. These relationships are extremely important, defining as they do the semantic nature of texts and providing essential keys to text structure.

Fortunately, such relationships have remained central in rhetorical theory from classical times to the present, from Aristotle's treatment of the enthymeme to the ideas advanced by contemporary theorists—including Christensen's (1965, 1966) notion of semantic levels of generality in supersentential discourse, Pitkin's (1969, 1977a, 1977b) theory of discourse "blocs" and hierarchies, Rodgers' (1966a, 1966b) idea of stadia in written

discourse, Becker's (1965, 1966) concept of functional semantic slots in paragraphs, and Larson's (1971) theory of linearity in the essay, to name but a few.

The centrality of supersentential relationships in texts is also evident in the recent work in text linguistics and discourse comprehension, from such early formal theories such as Harris' (1952) and Pike's (1954) to comprehensive theories such as Grimes' (1975) theory for delineating text structure, Frederikson's (1975) model for text representation, and Crothers' (1979) model for explicating the role of nonexplicit text inferences in text coherence.

Until recently, semantic relationships in extended units of discourse have been difficult to explore empirically, although Miles (1963, 1966), Ross (1970), Winterowd (1970), and Cummings, Herum, and Lybert (1971) obviously see the need for analyzing and describing written texts in terms of the explicit coherence markers contained in them. But it was only with the publication of Halliday and Hasan's Cohesion in English (1976) that a systematic and comprehensive method became available for analyzing coherence relationships in extended written discourse.

In the few minutes allotted to this presentation, we shall attend to three matters: (1) a brief introduction to the principles underlying Halliday and Hasan's method of analyzing cohesion in written discourse, (2) a summary of a study of the cohesive devices used by "good" and "poor" freshman writers, and (3) some suggestions for studying written discourse and discourse production from the standpoint of explicit cohesive relations.

Cohesion in Written Discourse

Trying to summarize briefly Halliday and Hasan's Cohesion in English is

a little like trying to illustrate the movements of our solar system with ten marbles on a two-dimensional surface: even the most important details must give way to the most general principles. Halliday and Hasan provide, not a theory of text structure, but a system for analyzing the cohesive relations which allow us to recognize a sequence of T-units as a complete text. For Halliday and Hasan, "the concept of cohesion is a semantic one; it refers to relations of meaning that exist within the text, and that define it as text" (p. 4). Obviously, then, the text is a "semantic unit" (p. 4). And at its most basic level, "cohesion is a semantic relation between an element in a text and some other element that is crucial to the interpretation of it" (p. 8). Central to Halliday and Hasan are the notions of "presupposing" and "presupposed" elements; and although "presupposing" and "presupposed" elements or items can co-occur in the same T-unit, it is the "cohesive tie" across T-unit boundaries which allow sequences of T-units to be realized as a single and complete coherent text. Halliday and Hasan's system for analyzing and classifying cohesive devices provides for five major classes of cohesive ties, nineteen subclasses, and numerous sub-subclasses. In the summary of a study of cohesion which follows this introduction, we will be concerned only with the five major classes—Reference, Substitution, Ellipsis, Conjunction, and Lexical Reiteration and Collocation—and their respective subclasses. Examples of these classes and subclasses appear on the handout distributed earlier. The first three of the major classes—Reference, Substitution, and Ellipsis—represent grammatical ties capable of reaching across T-unit boundaries; the fifth major class—Lexical Reiteration and Collocation—includes ties which are strictly lexical; and the fourth major class—Conjunction—contains ties which are lexical-grammatical. Three other important concepts need to be introduced here; these, too, are illustrated

on the handout distributed earlier. The first involves the distinction between "endophoric" and "exophoric" reference, the former including within-text references and the latter embracing references to presupposed items which lie outside the text itself. As far as Halliday and Hasan are concerned, "exophoric" reference "makes no contribution to the cohesion of a text" (p. 53). "Endophoric" reference—and this is the second concept—can be either "anaphoric," referring to some prior lexical item or items in the text, or "cataphoric," referring to some lexical item or items which follow in the text. The third concept—which might be called the "span of text" concept—allows cohesive ties to be classified according to the amount of text a given tie spans. Halliday and Hasan provide for four such "text-span" classes. If a cohesive tie simply links two adjacent T-units, it is an immediate tie. If a T-unit containing a presupposing item (say, for example, the pronominal it) of a cohesive tie is separated from the presupposed item by an intervening T-unit containing another instance (say, another it) of an item presupposing the same presupposed item, then the cohesive tie is a mediated one. If a presupposing item in one T-unit is separated from the presupposed item by one or more T-units containing no presupposing items relevant to the cohesive tie in question, then the tie is remote. Finally, a cohesive tie is said to be mediated and remote when the presupposing and the presupposed items are separated by at least two T-units, one of which contains another presupposing item relevant to the tie in question and one of which does not contain such an item.

Cohesion in "Good" and "Poor" Freshman Essays

Using procedures for analyzing and classifying cohesion in written discourse

outlined by Halliday and Hasan (pp. 333-355), we analyzed 10 University of Texas freshman essays written on the "changes in behavior" writing assignment used in the Miami University sentence-combining experiment (Morenberg, et al., 1978). Five "high" and five "low" quality essays were randomly selected from a much larger group which had been submitted to holistic evaluation. We anticipated that an analysis of cohesive ties would distinguish between the two essay sets. And it did, although time permits little more than a summary of the distinctions we found.

At the most general level of analysis, the two essay sets differ according to the mean number of words separating each instance of a cohesive tie. In the poorer essays, a cohesive tie of some type occurred once every 4.88 words; and in the better essays, a tie appeared once every 3.16 words, a difference in mean frequency in excess of 1.7 words. A comparable difference obtained for the mean number of cohesive ties per T-unit, with about 2.42 ties appearing in each T-unit in the poorer essays and about 5.17 ties appearing in each T-unit of the better essays. This cohesive density is similarly reflected in the total number of cohesive ties per 100 T-units in each essay set. In the poorer essays, 242.4 ties appeared per 100 T-units, but over twice that many, 516.8, occurred per 100 T-units in the "high" quality essays. Because the T-unit is used as a common denominator in several of the indices of cohesive density, it should be pointed out that the better essays contained T-units which were, on the average, 1.64 words longer than those in the poorer essays. By dividing the number of words in an essay set by the number of cohesive ties in that set, we arrive at another general index of cohesive density. In the better essays 31.66 % of all words contribute to explicit cohesive ties while only 20.46 % of the words in the poorer essays contribute to such ties.

The "span of text" classifications—immediate, remote, mediated, and mediated-remote—also help distinguish between the two essay sets. Both groups of writers used about the same relative percentage of remote ties, about 29.7 % for the poorer writers and about 26.9 % for the better writers. However, the better writers used a substantially higher relative percentage, about 41.6 % to 32.8 %, of immediate ties and a substantially smaller percentage, about 25.9 % compared to 36.7 %, of mediated-remote ties. Mediated ties appeared in the better essays with about eight times the frequency that they appeared in the poorer essays.

These percentages allow us to focus on some crucial differences between the two essay sets. The higher relative percentage of immediate cohesive ties in the better essays suggests, among other things, that the better writers tended to establish stronger cohesive bonds between individual T-units than did the poorer writers. This observation finds support in the analyses of reference and conjunction cohesion in the two essay sets. The analysis of reference cohesion shows that the better writers employed reference cohesion about twice as often, 84.1 times to 47.8 times, in a text span of 100 T-units as did the poorer writers. Because the largest difference in the occurrence of referential cohesion is located in the higher frequency in the better essays of personal pronouns (25.1 per 100 T-units as compared to 5.1 for the poorer writers) and because personal pronouns do not usually bridge intervening T-units, we can infer that the better writers paid considerably more attention to elaborating, in subsequent and adjacent T-units, lexical-semantic items introduced in a given T-unit. Also figuring importantly in the greater use of immediate cohesive ties is the frequency with which the better writers used conjunction to span and, therefore, link individual T-units. Conjunctive ties most often result in immediate cohesive ties

between T-units. It is not surprising, then, to find that the better writers used over three times as many conjunctive ties, 65.4 as compared to 20.4, per text span of 100 T-units as did the poorer writers. Neither is it surprising to discover that the better writers used all five types of conjunction while the poorer writers used only three. Thus it would appear that the better writers tended to extend the lexical-semantic domain established in a given T-unit considerably more often than did the poorer writers. One major effect of such lexical-semantic extensions is, of course, essay length; and the better writers wrote, on the average, 375.2 more words per essay than did the poorer writers.

As was pointed out earlier, the relative percentage of mediated ties is roughly eight times larger for the better essays. This finding further supports the notion of extending the lexical-semantic domain of a given T-unit. Mediated ties, in contrast to immediate ties, enable a writer to extend the lexical-semantic domain of a given T-unit into not merely an adjacent T-unit but into several succeeding T-units. Obviously, the better writers could do so with greater facility than could their poorer counterparts.

The higher relative percentage, 36.7 % as compared to 25.9 %, of mediated-remote ties in the poorer essays suggests another and important fundamental difference between the two groups of writers. This higher relative percentage of mediated-remote ties in the poorer essays tends to reinforce the suggestion that the poorer writers failed generally to elaborate and extend lexical-semantic domains through successive T-units and may well indicate that less "really" new information or semantic content was introduced during the course of a poorer essay than during the course of a better essay. That is to say, the poorer writers tended more toward reiteration of previously cued information than did the better writers. Indeed, in reading

the poorer essays one can not help but note a good deal of what might be called conceptual and lexical redundancy. Although from the standpoint of text cohesion a certain amount of redundancy is a virtue, the "redundancy" in the poorer essays seems to be a flaw, because the poorer writers simply failed to capitalize on occasions for creating immediate and mediated cohesive ties by supplying additional information at the point where it would be expected to appear.

The relative percentages of remote ties tends to reinforce this interpretation. Although the relative percentages do not differ markedly, 29.7 % for the poorer writers and 26.9 % for the better, they do suggest that the poorer writers tended more often (and the more often is here important) to reach back across several T-units in order to recover lexical-semantic items. This, coupled with the smaller relative percentage of immediate and mediated ties and the higher relative percentage of mediated-remote ties, may again suggest a greater reliance on a superficial redundancy of concept and lexical item in the "low" quality essays than in the "high" quality essays.

If time permitted, a much more detailed comparison of the two sets of freshman essays could be offered, a comparison examining the differences between the two sets along the dimensions of all five major classes and nineteen subclasses provided for in Halliday and Hasan's taxonomy. But enough of a comparison has been offered to indicate the usefulness of cohesion analysis in the study of student writing, a usefulness also suggested by a similar comparison conducted by Charles Cooper (Cooper, et al., 1979) and others of the writing of "good" and "poor" SUNY-Buffalo freshmen. The findings reported in the two studies also might have been compared with an eye to differences which might obtain for different ability-groups and

writing topics.

Directions for Further Research in Cohesion

If analyses of cohesion do, in fact, help us distinguish in rather specific ways between "good" and "poor" student writers, then such analyses hold a great deal of promise for the teaching of writing and suggest a number of research issues which, when properly investigated, can be brought to bear on classroom practice. What follows are a few possible directions research in this area might take. These directions can best be indicated by a series of questions which lend themselves to empirical investigation.

(1) From a developmental point of view, what distinctions between age-groups and between ability-groups within age-groups can be made on the basis of the frequencies and kinds of cohesive ties which appear in written discourse of different kinds?

(2) If such distinctions are found, do they parallel the development of syntactic complexity in written discourse?

(3) To what extent do the mode and the aim of discourse influence the frequencies and types of cohesive ties used by different writers?

(4) To what extent do the time allotted for writing and the subject matter influence the frequencies and types of cohesive ties used by different writers?

(5) Do the same frequencies and types of cohesive ties obtain across repeated measures of writing when the subject matter, the aim, and the mode of discourse are similarly controlled

across both measures?

(7) How much of the variance in holistic scores can be predicted by the values computed for the various indices of cohesive density?

(8) To what extent do the revision practices of "skilled" and "unskilled" writers affect text cohesion?

(9) Can particular revisions be classified according to Halliday and Hasan's taxonomy of cohesive ties?

(10) What effects do the kinds and frequencies of cohesive ties have on the readability of particular kinds of written texts?

(11) To what extent are choices made during the composing process governed by considerations of text cohesion?

(12) Can analyses of cohesive ties enable us to identify objectively such discourse features as discourse "blocs" and stadia, semantic levels of generality, text hierarchies, and linear and non-linear structures?

Such a list could, of course, go on indefinitely, but enough of the potential that cohesion analysis holds has been suggested to indicate that it might well provide quantitative researchers with, if not a comet, at least a flame by which to light their tapers for years to come. It is hoped, however, that cohesion will never be promoted with the lack of respect for its limitations that transformational grammar was by certain of its over-zealous proponents. Cohesion studies have the potential to lead us into areas of investigation previously inaccessible, but in leading us to those areas, it will probably never give composition teachers the panacea for teaching writing that so many sought through applications of transformational grammar in the composition classroom.

References

Becker, A.L. Symposium on the paragraph. College Composition and Communication, 1966, 17, 72-80.

Becker, A.L. A tagmemic approach to paragraph analysis. College Composition and Communication, 1965, 16, 237-242.

Chomsky, N. Syntactic structures. The Hague: Mouton, 1957.

Christensen, F. A generative rhetoric of the paragraph. College Composition and Communication, 1965, 16, 144-156.

Christensen, F. Symposium on the paragraph. College Composition and Communication, 1966, 17, 60-66.

Cooper, C. Writing abilities of regularly-admitted freshmen at SUNY-Buffalo. Unpublished paper. 1979.

Crothers, E.J. Paragraph structure inference. Norwood, N.J.: Ablex, 1979.

Frederiksen, C.H. Representing logical and semantic structure of knowledge acquired from discourse. Cognitive Psychology, 1975, 7, 371-458.

Grimes, J.E. The thread of discourse. The Hague: Mouton, 1975.

Halliday, M.A.K., and Hasan, R. Cohesion in English. London: Longman, 1976.

Harris, Z.S. Discourse analysis: a sample text. Language, 1952, 28, 474-494.

Hunt, K.W. Grammatical structures written at three grade levels. Research Report, No. 3. Urbana, Illinois: NCTE, 1965.

Larson, R.L. Toward a linear rhetoric of the essay. College Composition and Communication, 1971, 22, 140-146.

Mellon, J. Transformational sentence-combining: a method for enhancing the development of syntactic fluency in English composition. Research Report, No. 10. Urbana, Illinois: NCTE, 1969.

Miles, J. Symposium on the paragraph. College Composition and Communication, 1966, 17, 80-82.

Miles, J. What we compose. College Composition and Communication, 1963, 14, 146-154.

Morenberg, M., Daiker, D., and Kerek, A. Sentence combining at the college level: an experimental study. Research in the Teaching of English, 1978, 12, 245-256.

Pike, K.L. Language in relation to a unified theory of the structure of human behavior, Part 1. Glendale, California: Summer Institute of Linguistics, 1954.

Pitkin, W., Jr. Discourse blocs. College Composition and Communication, 1969, 20, 138-148.

Pitkin, W., Jr. Hierarchies and the discourse hierarchy. College English, 1977a, 38, 649-659.

Pitkin, W., Jr. X/y: some basic strategies of discourse. College English, 1977b, 38, 660-672.

Rodgers, P.C., Jr. A discourse-centered rhetoric of the paragraph. College Composition and Communication, 1966a, 17, 2-11.

Rodgers, P.C., Jr. Symposium on the paragraph. College Composition and Communication, 1966b, 17, 72-80.

Ross, D., Jr. Composition as stylistic features. Style, 1970, 4, 1-8.

Winterowd, W.R. The grammar of coherence. College English, 1970, 31, 828-835.

Stephen P. Witte, University of Texas at Austin
CCCC Handout
March, 1980

Cohesion in English: Examples of Types

REFERENCE: Pronominal

At home, my father is himself. //
He relaxes and acts in his normal manner. //

REFERENCE: Demonstratives and Definite Articles

Two coats were hanging in the closet. //
This coat fell on the floor. //

Humans have many needs, both physical and intangible. //
It is easy to see the physical needs such as food and shelter. //

REFERENCE: Comparatives

The males ruled the women; //
the women obeyed and formed themselves into the mold the males constructed for them. //
But today it is different. //

SUBSTITUTION: Nominal

I have another behavior for the classroom. //
That one society . . . deems appropriate for the classroom. //

SUBSTITUTION: Verbal

A father may expect his son to develop interest and skill in playing football; //
but the son may not do so. //

SUBSTITUTION: Clausal

Do all people change their behaviors this easily? //
It seems so.

ELLIPSIS: Nominal

Among most teenagers we can observe three types of behavioral patterns. //
These might be called the good, the bad, and the ugly. //

ELLIPSIS: Verbal

Do these changes in roles point to insincerity within us? //
To a dgree, they do.

ELLIPSIS: Clausal

Can role-playing be reduced? //

- Yes, if our society becomes more liberal. //

CONJUNCTION: Additive

The college student who wore faded blue jeans now wears three-piece suits, //
and he feels the need to be approved of and accepted by his new associates. //

CONJUNCTION: Adversative

People can usually be themselves when they are with friends. //
However, people who are attempting to make friends often feel a need to impress. //

CONJUNCTION: Temporal

A friend of mine went to an out-of-state college, //
and before she left, she expressed her feelings about playing roles to win new friends. //

CONJUNCTION: Continuative

No one wants to be rejected, //
and to prevent rejection we, of course, change our behavior often. //

LEXICAL REITERATION: Same Item

Drinking is typical of college freshmen. . . . //
The drinking is usually accompanied by obscene talk. //

LEXICAL REITERATION: Synonym or near-synonym

Drinking is typical of college freshmen. . . . //
This boozing is usually accompanied by obscene talk. //

LEXICAL REITERATION: Superordinate

Drinking is typical of college freshmen. . . . //
The behavior is usually accompanied by obscene talk. //

LEXICAL REITERATION: General item

Drinking is typical of college freshmen. . . . //
The thing is usually accompanied by obscene talk. //

LEXICAL COLLOCATION

When we are not with our friends, we like being pampered by our parents, //
but when we are with our friends, we hate it. //

Exophoric Reference

In the first part of the essay, we will examine the effects of mass media on human behavior. //

Endophoric Reference

John bought an apple. //
He bought it yesterday. //

Anaphoric Reference

John bought an apple. //
He bought it yesterday. //

Cataphoric Reference

That was good. //
It was the best meal I've ever eaten. //

Immediate Cohesive Tie

John bought an apple. //
He ate it today. //

Remote Cohesive Tie

John bought an apple. //
He also bought some oranges. //
And he bought some bread. //
But he didn't eat the apple until yesterday. //

Mediated Cohesive Tie

John bought an apple, //
and he carried it home in a brown bag. //
But he didn't eat it until this morning.

Mediated-Remote Cohesive Tie

John bought an apple, //
and he carried it home in a paper bag. //
At home he played with Homer for a short time, //
and then he ate the apple. //